Adapting an Agent-Based Model of Socio-Technical Systems to Analyze Security Failures

Christine Cunningham
MIT Lincoln Laboratory
Lexington, MA
Telephone: (781) 981–5500
Christine.Cunningham@ll.mit.edu

Antonio Roque
MIT Lincoln Laboratory
Lexington, MA
Telephone: (781) 981–5500
Antonio.Roque@ll.mit.edu

I. Introduction

This paper works towards developing agent-based models of socio-technical systems to study security failures. These models could contribute to multi-component simulations such as cyber range events, where they would generate network traffic[8][9]. Agent-based modeling[10] involves autonomous and proactive programs which communicate peer-to-peer. Socio-technical system approaches involve models of humans, their organizations, the tools they use, and the interaction between all of these[11]. Agent-based models of socio-technical systems have been applied in the context of air traffic systems of air traffic controllers and pilots[11], economic production/consumption networks[12], and more.

The components of our model need to be justified in terms of theory or data, preferably both. As a starting point we adapt an agent-based model of socio-technical systems developed by Crowder et al.[13], which is based on concepts from industrial psychology, on data collections, and on discussions with subject matter experts.

II. Blackout Scenario

We are particularly interested in the security of electrical power grids, which are considered one of a nations's *Critical Infrastructure and Key Assets*, whose effectiveness and security are vital to maintain[14], and which are a potential target for attack[15].

Finding a good historical example relevant to critical infrastructure is challenging because of the need for a scenario that is well-documented and realistic. Unfortunately, security and privacy concerns make finding such information difficult. Our solution is to identify a scenario from a general system failure case. System failures are disruptions of normal functions, and security failures can been seen as *intentional* system disruptions[16]. We therefore adapt a past system failure case whose causes could plausibly have been intentional and computer-related.

The scenario we identified is the 2003 Northeast Blackout, which was the largest blackout in North American history. It affected 50 million people (including over 20 million in the New York City and 8 million in the Toronto metropolitan areas) and cost an estimated 6 billion dollars[17], revealing vulnerabilities in the infrastructure and management of the electrical power grid. Unless otherwise stated, the description details in Section II are taken from the North American Electric Reliability Corporation's "Final Report on the August 14, 2003 Blackout in the United States and Canada."[18]

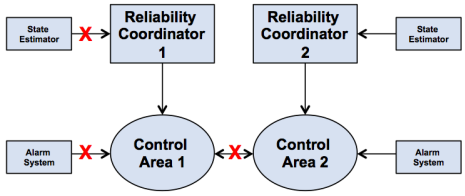

Fig. 1. Organizational structure of relevant entities in the 2003 Northeast Blackout.

Figure 1 shows the organizational structure of the socio-technical elements contributing to the 2003 Northeast Blackout.

Reliability Coordinators (RCs) cover multi-state regions; they are responsible for monitoring and coordinating their multiple **Control Areas** (CAs) as well as the CAs of their neighbors. RCs must provide yearly, monthly, and daily energy consumption predictions, as well as contingency analyses for managing electrical flow during unanticipated situations. One of the tools that RCs use is a **state estimator**, which enables these contingency analyses.

The 2003 Northeast Blackout was not the result of a computer attack, but each of the socio-technical system's failures could have been caused by a computer attack without changing the way that the system reacted. What we call our **Blackout scenario** therefore is an extraction of the socio-technical security vulnerabilities displayed during the 2003 Northeast Blackout.

III. Agent-Based Model

A. Original Model

Given our need for realistic user agents when implementing the scenario, we next sought out an appropriate user model. We chose an agent-based model developed by Crowder et al.[13] which applied socio-technical systems theory to modeling work teams. Crowder et al. developed their model with concepts from psychology, management, and computer modeling,

as well as with quantitative and qualitative data collected from multidisciplinary engineering teams. Their model describes how a task's requirements cause team members to communicate among themselves, and the cognitive mechanisms that integrate the results of that communication.

The Crowder et al. model includes a Task Workflow Model which describes the steps required to complete a task, dependencies between the steps, the difficulty of each step, and the team member responsible for completing the task. Additionally, an Agent Model with components such as *Trust* and *Shared Mental Models* uses a set of equations to describe the interaction of those components while performing a task. The model produces a set of completion and working times for performing the task, as well as a measure of task quality. Finally, a Communication Model describes the way that agents request information as needed, to increase their ability to complete a task step. More details about these models is contained in Section III-B where we describe adapting them to our scenario.

B. Adapting the Model

1) Task Workflow Model: Figure 2 shows the **Task Workflow Model**. It includes 4 tasks, each which has an agent assigned to it (either CA1 or RC1) and a *Task Difficulty* value between 0 and 5, following Crowder et al.'s use of semantic labels on that scale.

To determine the Task Difficulty, we produced qualitative descriptions of the information available to the agents, as well as of the stakes involved. As the scenario progresses, more information is available to the agents about the nature of the problem; in that way the tasks are easier. As the scenario progresses the stakes are higher, though, so in that way the tasks are harder. This analysis produced a quantitative estimate based on the dynamic between these qualitative factors .

2) Agent and Communication Models: The Crowder et al. model was developed from an engineering domain that took weeks and months to perform, rather than the shorter timeframe involved in the Blackout scenario. This led us to make several changes to our **Agent Model**.

The Crowder et al. Agent Model produced several outputs: the *Completion Time* tracks how long it takes an agent to finish a particular task; these are combined from all tasks and agents to produce a Total Completion Time. The *Working Time* tracks how much time the individual agent spends working on a task; these are combined from all tasks and agents to produce a Total Completion Time. The *Quality* describes the degree of excellence of the task once finished; these are combined from all tasks and agents to produce a Total Quality.

In our Blackout scenario, the agents were working under strict time constraints. They had a short amount of time to decide how to resolve problems they encountered, and at the end of time they had to address the problem, such as dropping power or shifting loads, but even if the operators dropped power, they might not drop enough. Doing nothing before time ran out was one way of handling the task, though in our scenario this was always the wrong decision. Because task completion time was a constraint, we removed the *Completion Time* and *Working Time* components, as well as the *Availability*, *Learning Time*, and *Response Rate* components which likewise were dependent on longer time-scales.

Next, we considered the *Shared Mental Models*, *Motivation*, and *Communication Frequency* components. The main distinctions between those components was that Shared Mental Models did not directly affect Quality, and Competency affected Communication but Motivation did not. However, the distinction between these components is more important in Crowder et al.'s use case than in ours: for example, they may be interested in changing these values to determine whether it is more cost-effective to invest in increasing team Motivation, or team Shared Mental Models. Furthermore in the Crowder et al. model, the equations that drove the algorithms behind these components contained numerous coefficients derived from regression analyses of Likert-scale questionnaire data taken from their engineering domain. To limit dependence on these domain-specific coefficients, we therefore merged the Shared Mental Models and Quality components into the Competency component, and included a normally (Gaussian) distributed value with a standard deviation of 1 in the equation for Competency as a way of partially substituting for their effect. Our final Agent Model is shown in Figure 3.

Crowder et al.'s version of *Trust* is an input value, modified by the success or failure of the agent's previous interactions with a team member on longer-lasting tasks. Because our tasks and time scales are different, our version of Trust is an input value on the 0-5 scale to produce the base trust τ_b. This base trust will be modified by a normally (Gaussian) distributed value v_τ with a standard deviation of 1 and a mean determined by the experiment settings as described in Section IV to produce a working trust τ_w.

$$\tau_w = \tau_b + v_\tau \tag{1}$$

Crowder et al.'s version of *Competency* is an input value, which is increased by interactions with other team members. Our Blackout scenario agents also have a base input value C_b modified by a normally distributed value v_C (which also has a standard deviation of 1 and a mean determined by the experiment settings). This may also be modified by an increase in competency due to interactions with other team members δ_C to produce a working competency C_w.

$$C_w = C_b + v_C + \delta_C \tag{2}$$

We use Crowder et al.'s equation for δ_C and similarly cap the possible Competency gain to 0.3.

Crowder et al.'s model assumes that a receiving agent C_r will continue to seek communications with other team members C_p who are providing information (thereby increasing Competency) until the agent has a Competency sufficient for the task difficulty. In the Blackout scenario, the CA1 agent has a chance to gain competency from the CA2 agent, but does not seek out further competency gain due to the Task time scale and absence of other agents to refer to. Therefore although our

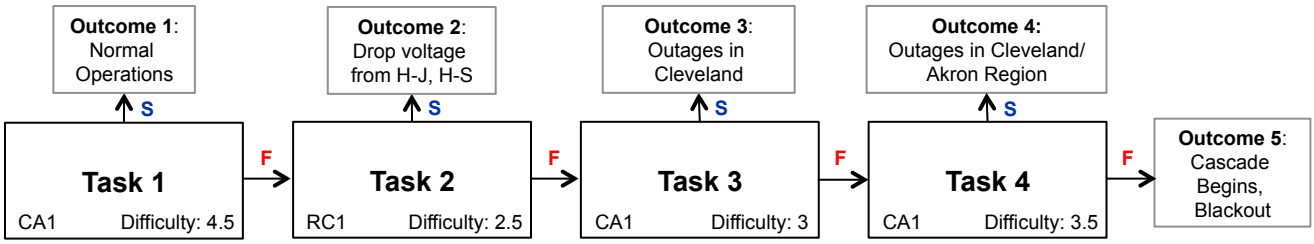

Fig. 2. Blackout Scenario Task Workflow Model. S=Success, F=Failure

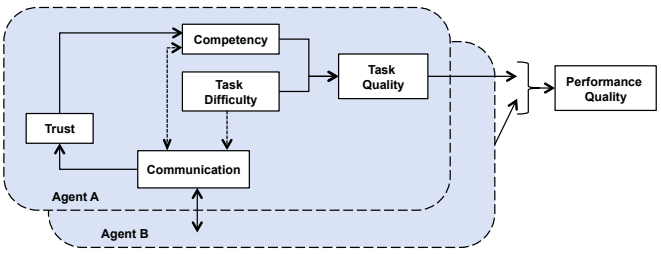

Fig. 3. Adapted Agent Model.

Competency can in principle interact with Communication, in this scenario's tasks it does not.

In our scenario, for a task n the *Task Quality* Q_n is a (0,1) value describing whether the voltage problems were *completely* resolved, because partial solutions did not stop the process leading towards blackout. This is determined by comparing the agent's working Competency to the Task Difficulty D_n. If $C_w \geq D_n$, then the Task Quality is 1 (success); otherwise it is 0 (failure).

The overall Performance Quality Q_P then expresses the performance on the i tasks of the Blackout scenario as an integer between 0 (for blackout) to 5 (for best outcome).

$$Q_P = \frac{5}{4}(4 - \sum_{n=1}^{i}(1 - Q_n)) \quad (3)$$

Finally, our **Communication Model** is shown in Figure 1. Communications are tied to the workflow model: each Task defines a communication that occurs between agents as part of their involvement in the system.

IV. Simulations

A. Baseline Simulation

We ran 12,500 iterations of the Blackout scenario to build a baseline. We ensured an iteration through the parameter space of all possible inputs by cycling through the base competency settings as shown in Table I. In this way we are sure to explore all possible combinations of team competencies, rather than be tied to a representation in which the teams are all of mid-level competency or high-level competency. Having explored the parameter space through 12,500 iterations, we then reduced the number of iterations for subsequent experiments.

TABLE I
Experiment Settings

	CA1 **Trust** in CA2	CA1 base **Competency**	CA2 base **Competency**	RC1 base **Competency**
1	VH (4.5)	VH (4.5)	VH (4.5)	VH (4.5)
2	VH (4.5)	VH (4.5)	VH (4.5)	H (3.5)
3	VH (4.5)	VH (4.5)	VH (4.5)	M (2.5)
...	...	...	...	...
625	VL (0.5)	VL (0.5)	VL (0.5)	VL (0.5)

TABLE II
Experiment Simulations

	Baseline	**Delegation 1**	**Delegation 2**
Total Iterations	12500	3125	3125
Outcome 1	1273 (10.2%)	314 (10.0%)	630 (20.2%)
Outcome 2	4495 (36.0%)	1125 (36.0%)	1000 (32.0%)
Outcome 3	3082 (24.7%)	757 (24.2%)	829 (28.5%)
Outcome 4	0 (0%)	377 (12.1%)	243 (7.8%)
Outcome 5	3650 (29.2%)	552 (17.7%)	360 (11.5%)
Total Blackout	3650 (29.2%)	552 (17.7%)	360 (11.5%)
Total Non-Blackout	8850 (70.8%)	2573 (82.3%)	2765 (88.5%)

Table II shows the results, along with the number of times each outcome occurred. The outcomes described here are those shown in Figure 2 and described in Section III-B1: Outcome 1 is *Normal Operations*, Outcome 2 is *Drop Voltage from H-J and H-S lines*, Outcome 3 is *Outages in Cleveland*, Outcome 4 is *Outages in Cleveland/Akron region*, and Outcome 5 is *Cascade Begins, Blackout*.

As a metric for determining the efficiency of the agent system, we used the percentage of iterations in which $Q_p > 0$. This is the percentage of Non-Blackouts, the number of times the scenario ended in Outcomes 1-4 (i.e. was resolved by dropping voltage from lines or regions, even if resulting in smaller local outages) instead of reaching Outcome 5, (i.e. ended in a cascade leading to a major blackout as in the 2003 Northeast Blackout.) The total and percentage of Non-Blackouts is shown in the last row of Table II.

B. Delegation Experiments

As a first experiment, we explored a policy which might improve upon the situation where Outcome 4 never occurs. We implemented an agent team policy in which, upon reaching Task 4, that task is delegated to another agent. This was effected by replacing the CA1 agent with a CA1 agent with

a different Competency. The rationale for this is that CA organizations are actually made up of numerous agents. We were previously assuming that a single CA agent would handle every task, but it is equally reasonable to assume that a CA organization would have a set of agents, and that the organization's policy would be to randomly assign tasks that have been received. We hypothesized that this policy would improve the Total Non-Blackout metric. 3125 simulation iterations using this delegation policy produced the data summarized in the Delegation 1 column of Table II; we found that the total number of non-blackouts differed from the total number in the baseline data to a statistically significant extent with a p-value$<$0.0003.

However, this is slightly unrealistic because it assumes that the CA team has a reliable way of knowing that they were in a task that should be delegated to another agent. As a second experiment, we therefore hypothesized that delegating *each* of the tasks in the scenario (instead of only Task 4) to different agents would significantly increase performance. We implemented this as instantiating a new agent with a new competency rating for each task in Figure 2. 3125 simulation iterations using this new delegation policy produced the data summarized in the Delegation 2 column of Table II; we found that the total number of non-blackouts differed from the total number in the baseline data to a statistically significant extent with a p-value$<$0.0003. Note also the increase in the best possible outcome of Outcome 1, and the decrease in the two worst Non-Blackout outcomes of Outcomes 4 and 5.

ACKNOWLEDGMENT

This work is sponsored by the Test Resource Management Center under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government. This work was initiated while Christine Cunningham was affiliated with Williams College.

BIOGRAPHIES

Christine Cunningham is a member of the Cyber System Assessments Group at MIT Lincoln Laboratory. She graduated from Williams College in 2015 with a BA degree in computer science.

Antonio Roque has been a member of MIT Lincoln Laboratory since 2013. He works on the research and development of high-fidelity traffic generators, as well as on methodology for cyber security assessments. He received a PhD in Computer Science from the University of Southern California.

REFERENCES

[1] L. F. Cranor, "A framework for reasoning about the human in the loop," in *Proceedings of the 1st Conference on Usability, Psychology, and Security*. Berkeley, CA, USA: USENIX Association, 2008, pp. 1:1–1:15.

[2] A. E. Howe, I. Ray, M. Roberts, M. Urbanska, and Z. Byrne, "The psychology of security for the home computer user," in *IEEE Symposium on Security and Privacy*, 2012, pp. 209–223.

[3] T.-F. Yen, V. Heorhiadi, A. Oprea, M. K. Reiter, and A. Juels, "An epidemiological study of malware encounters in a large enterprise," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. pp 1117–1130.

[4] The Smart Grid Interoperability Panel - Cyber Security Working Group, "Guidelines for smart grid cyber security: Vol. 1, Smart grid cyber security strategy, architecture, and high-level requirements," National Institute of Standards and Technology, Tech. Rep., August 2010.

[5] V. Kothari, J. Blythe, S. Smith, and R. Koppel, "Agent-based modeling of user circumvention of security," in *Proceedings of the 1st International Workshop on Agents and CyberSecurity*. ACM, 2014.

[6] J. Blythe, A. Botello, J. Sutton, D. Mazzocco, J. Lin, M. Spraragen, and M. Zyda, "Testing cyber security with simulated humans," in *Proceedings of the Twenty-Third Innovative Applications of Artificial Intelligence Conference*, 2011.

[7] C. V. Wright, C. Connelly, T. Braje, J. C. Rabek, L. M. Rossey, and R. K. Cunningham, "Generating client workloads and high-fidelity network traffic for controllable, repeatable experiments in computer security," in *Proceedings of Recent advances in intrusion detection*, 2010.

[8] L. M. Rossey, R. K. Cunningham, D. J. Fried, J. C. Rabek, R. P. Lippmann, J. W. Haines, and M. A. Zissman, "Lariat: Lincoln adaptable real-time information assurance testbed," in *Aerospace Conference Proceedings*, 2002.

[9] S. K. Damodaran and J. M. Couretas, "Cyber modeling & simulation for cyber-range events," in *Proceedings of Summer Computer Simulation Conference*, 2015.

[10] S. F. Railsback and V. Grimm, *Agent-Based and Individual-Based Modeling*. Princeton University Press, 2012.

[11] K. H. van Dam, I. Nikolic, and Z. Lukszo, Eds., *Agent-based modelling of socio-technical systems*. Springer Science & Business Media, 2013, vol. 9.

[12] A. P. Shaw and A. R. Pritchett, "Agent-based modeling and simulation of socio-technical systems," *Organizational Simulation*, pp. 323–367, 2005.

[13] R. M. Crowder, M. A. Robinson, H. P. Hughes, and Y.-W. Sim, "The development of an agent-based modeling framework for simulating engineering team work," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 42, no. 6, pp. 1425–1439, 2012.

[14] J. Moteff and P. Parfomak, *Critical Infrastructure and Key Assets: Definition and Identification*. Congressional Research Service, October 1 2004.

[15] P. Shakarian, J. Shakarian, and A. Ruef, *Introduction to cyber-warfare: a multidisciplinary approach*. Elsevier, 2013.

[16] W. Young and N. Leveson, "Systems thinking for safety and security," in *Proceedings of the 29th Annual Computer Security Applications Conference*, 2013, pp. 1–8.

[17] J. Minkel, "The 2003 Northeast Blackout - five years later," *Scientific American*, August 2008.

[18] US-Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and recommendations," http://www.nerc.com/pa/rrm/ea/Pages/Blackout-August-2003.aspx, North American Electric Reliability Corporation, Tech. Rep., April 2004.

[19] F. Bellifemine, G. Caire, and D. Greenwood, *Developing Multi-Agent Systems with JADE*. Wiley, 2007.

[20] J. D. Faus and F. Grimaldo, "Infraworld, a multi-agent based framework to assist in civil infrastructure collaborative design," in *11th International Conference on Autonomous Agents and Multiagent Systems*, 2012.

[21] Z. M. Ibrahim, L. F. de la Cruz, A. Stringaris, R. Goodman, M. Luck, and R. J. Dobson, "A multi-agent platform for automating the collection of patient-provided clinical feedback," in *2015 International Conference on Autonomous Agents and Multiagent Systems*, 2015.

[22] T. Michalak, J. Sroka, T. Rahwan, M. Wooldridge, P. McBurney, and N. R. Jennings, "A distributed algorithm for anytime coalition structure generation," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1, pp. 1007-1014. International Foundation for Autonomous Agents and Multiagent Systems*, 2010.